

Dynamic changes of pausing in triadic conversations

Dorottya Gyarmathy, Valéria Krepsz, Anna Huszár and Viktória Horváth
Hungarian Research Centre for Linguistics, Budapest, Hungary

Abstract

Pausing in conversation has several roles from speech planning to managing turn-takings (TTs). However, less is known about the dynamic changes of pauses over time or with regard to the turn-taking system. The frequency and the duration of silent and filled pauses (SPs and FPs) as well as shared silences was analyzed in 20 triadic Hungarian conversations using dynamic frames (altogether more than 7700 items). Data showed that the frequency of silent and FPs decreased over time across conversations. As opposite, shared silences were found to be the most frequent in the last sections of conversations. However, the duration of the pauses did not change over time across conversation—it may be influenced by other factors. We found that the SPs containing audible breathing were longer than other SPs. The SPs were less frequent before turn-takings than in other positions. However, their duration was not affected by the turn-taking system.

Introduction

Pauses have many kinds of roles in communication (e.g. respiration, cognitive load, production problems), both in speech production and perception next to boundary marking. So not every pause necessarily behaves as TRP (transition relevance place, which is defined as timing when the current speaker's turn can be completed and other participants are able to take the turn, cf. Sacks, Schegloff, & Jefferson, 1974), several factors can affect its appearance, frequency and duration as well. Local and Kelly (1986) investigated two different kinds of pauses: 1. 'trail-off silences' (a possible point for switching the role of the speaker) and 2. 'holding silences' (the speaker keeps the floor, it serves as an inhalation point or as a rhetorical tool). In case of trail-off pauses they found open glottis, out-breath, vowel centralization, and diminished loudness and tempo, preceded the pause. In the case of holding silence, they found closed glottis and no final lengthening preceding the pause. Levelt (1989) also differentiated types of silences according to their position and function. The speakers' tempo increased in the vicinity of syntactic boundaries to keep the floor and the rights of speaking, however they slow down and take a pause in the next phrase (Schegloff, 1996, Eggins & Slade, 1997).

The analysis of breathing in dyadic conversations corroborated that the speakers coordinate breathing to turn-takings (TTs). Inhalations inside a turn were shorter than when starting a new turn, suggesting that participants also adapt their breathing to hold turns (Rochet-Capellan & Fuchs, 2014). Inbreaths were analyzed in question-answer sequences in Dutch conversations, and they were found to be more frequent preceding long answers than short answers (Torreira, Bögels, & Levinson, 2015).

Filled pauses (FPs) also have several functions in the organization of the TT system as well. FPs may have pragmatic functions as indicators of the Feeling-of-Another's-Knowing in a dialogue (Brennan & Williams, 1995), or as turn-holders (Stenström, 1994). Therefore, some works described FPs as an interactional phenomenon (Levinson, 1983, Clark, 1994). FPs mark for the listeners that the next utterance will be more complex and the speaker needs more time for speech planning. Swerts (1998) found that FPs after stronger breaks tend to occur phrase-initially, whereas the majority of the FPs after weak boundaries are in phrase-internal position. The type and the position of FPs showed connection: 'um' was found to be more frequent at turn-initial position than 'uh', while 'uh' occurred rather at turn-medial position. Another study corroborated that FPs are often used to initiate the speaker's turn. In addition, when a speaker is confronted with unsuccessful answers in the course of the dialog, hesitations may also stand for marking his/her embarrassment and wish to close the dialog (Vasilescu, Rosset, & Adda-Decker, 2010). Isolated FPs occurred more frequently within their host unit than between clauses in English and French as well (Crible, Degand, & Gilquin, 2017). The FPs were also analyzed with regard to TTs from the Columbia Games Corpus (Benus, 2009). 33% of all FPs were in turn-initial position; so, FPs are linked to TT because these peripheral positions suggest several floor-management functions. FPs in this pre-start function allows the speaker some time for planning and the listener for tuning in.

The aim of the present study was to analyze the silent and filled pauses with regard to their position in the conversations. The main question was, how does pausing change across conversations? Which part of the conversation does contain the least pause or the shortest shared silences (ShS)?

Our hypotheses are the following:

1. We assume that during the conversations, the frequency and duration of silent, filled, and shared pauses decreases due to the accustoming and synchronization of the speakers. At the end of the conversations, an increase would be observed, as the participants run out of the topic of conversation and intend to close the communication event.

2. Silent pauses (SPs) are less frequent and shorter near to TTs, while more frequent and longer further from TTs.

3. FPs would occur less frequently in the position near before to the TTs than further from them. In addition, FPs would be more frequent and longer after turn-taking in turn-initial position (cf. Swerts, 1998, Benus, 2009).

Material and method

20 conversations were selected for the present study from the Hungarian Spontaneous Speech Database (BEA, cf. Neuberger et al., 2014) prepared in the phonetic lab of the Hungarian Research Centre for Linguistics. The BEA database consists of 460 recordings, which contain 7 different speech tasks, for example reading sentences and text, narratives. The conversation task is the 5th task in the whole recording. Three people participate in each conversation: the fieldworker1 (Fw1), the experimental speaker (S) and fieldworker2 (Fw2). The conversations are seminatural: the participants have no time for preparation, the first topic is given by the Fw1, but further topics are not fixed—the speech planning processes and the organization of the conversation are spontaneous. The two fieldworkers were the same people in each conversation (two female speakers, linguists, colleagues, 27–38 years old during conducting the database), while S changes across conversations (aged between 20–45 years). The conversations are about 18 mins long on average (8.5–23.5 min), the 20 conversations took almost 6 hours. The annotation of the material was carried out manually using Praat (Boersma & Weenink, 2018) by two trained annotators. The value of the inner-annotator agreement was 95%. In the case of disagreement, a third senior annotator checked the problematic parts and helped to decide. The annotation includes the level of interpausal units of the 3 speakers, the SPs and the hesitations as well. Furthermore, TTs, overlapping speech, backchannel responses were annotated in additional tiers (Horváth et al., 2019).

The patterns of the (silent and filled) pauses as well as shared silences were analyzed: i) frequency, ii) duration iii) types iv) audible breathings in SPs. Silent pauses and shared silences were differentiated based on their position: SPs were defined within a

speaker's utterance, while ShSs were defined between the different speaker's units, when no one was speaking. The analysis was carried out using a dynamic approach: how these patterns change i) over time across conversations ii) near and further from TTs? The changes over time were analyzed using the following method: each conversation was split into 5 equal parts based on their duration automatically by a Praat script. For example, a 15-minute long conversation was cut up into five 3-minute long subsections (0–20%, 21–40% etc.). With this method, we can eliminate the unequal durations of the conversations, and the occurrences of the given parameter can be comparable. The connection between pauses and TTs was analyzed with the following method (see Figure 1). The distance between pauses and the nearest TT was extracted automatically using a Praat script. The pauses and in most of the cases the TTs are not a point extend phenomena; therefore, we calculated with the centers of the intervals. Based on the distance values of the pauses from the nearest TTs, the pauses were split into four groups according to two parameters: 1) nearer or further from TTs 2) before or after TTs. The border of closure vicinity was determined at 5 s based on the length of turns and the context.

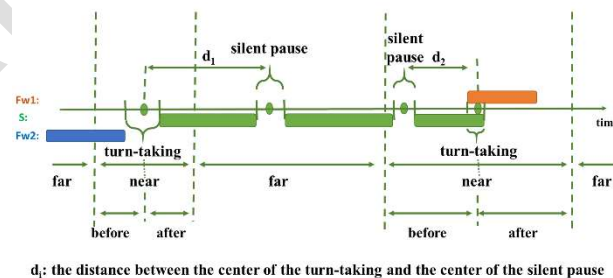


Figure 1. Method of the analysis of pauses near and further from TTs

The duration of the pauses was analyzed with linear mixed models in the R program (R Core Team, 2018) with the lme4 package (Bates et al., 2015), and the p values were obtained by Satterthwaite approximation (lmerTest package, ANOVA function, Kuznetsova, Brockhoff, & Christensen, 2015). The independent factors were the duration of the pauses, while the dependent factors were the five-partitions of the conversations. For each parameter, a random intercept and slope model were used (with the speaker as a random factor for each variable) and compared to the two models. There were no significant differences between the models, and because of the lower AIC (Akaike, 1973) values, the random intercept models were used during the analyses. The frequency of the pauses were analysed

with Friedman Test (R Core Team, 2018): the dependent variable was the frequency of the pauses and the independent variable was the position of the pauses (near before, near after, further before, further after TT).

Results

Silent pauses

5881 SPs occurred in the 20 recordings. The mean frequency was 18.03 SPs/min ($SD = 4.76$). The frequency of SPs was analyzed in the 5 equal parts of the conversations. Results showed that SPs' occurrence was affected by their position in the conversation: $\chi^2(4) = 18.025$, $p = 0.001$ (Figure 2). They occurred least frequently in the first part of the conversation, then their frequency increased in the middle sections, while decreased again in the last section of the conversations.

The mean duration of the SPs was 431 ms ($SD = 336$ ms). The duration of SPs was also analyzed with regard to their position in the conversation. Data showed that there was no significant difference between the sections of the conversation in the duration of SPs. SPs were also analyzed with regard to their breathiness. 35% of the SPs contained audible breathing. The SPs with audible breathing were significantly longer than pauses without audible breathing [$F(1, 5781) = 248.625$, $p < 0.001$], irrespectively of the participant's role (Figure 3).

The mean duration of SPs with audible breathing were 577 ms ($SD = 305$ ms), without audible breathing were 354 ms ($SD = 326$ ms).

Filled pauses

A total of 1240 FPs occurred in the 20 recordings. The mean frequency of the FPs was 3.77 item/min ($SD = 2.31$). The dynamic change in the frequency data was analyzed in the 5 equal parts of the conversations. FPs occurred the least frequently in the last section ($mean = 3.5$ item/min), while the most frequently in the 2nd section ($mean = 3.87$ item/min). However, the difference was not significant between the sections. The type of the FPs was analyzed. 57% of the FPs occurred as a monophthong schwa, while 35% realized as a nasal consonant. The ratio of diphthongs (like [əɪ] or [əʊ]) was altogether less than 10%. The duration of FPs significantly differed from their forms ($F(1239, 4) = 31.439$, $p < 0.001$): the more sounds the FP involved the longer duration it had (e.g. the duration of the swa form ([ə]) was on average 306 ms, while the average duration of [əhm] was 630 ms).

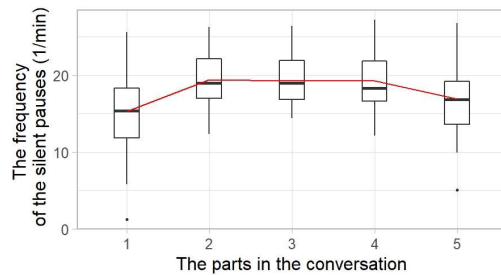


Figure 2. The frequency of the SPs in the 5 equal parts of the conversations (red line represents the means while black line on the boxes represents the medians).

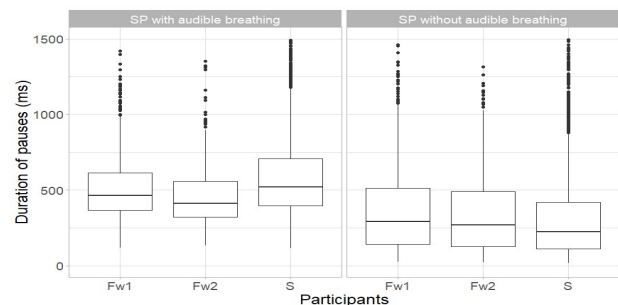


Figure 3. The duration of the SP with regard to breathiness and participant's role.

Shared silences

628 shared silences occurred in the 20 conversations with a mean duration of 510 ms ($SD = 513$ ms). The frequency of silences was 1.98 item/min on average ($SD = 1.38$). The frequency was also analyzed with regard to the changes over time across conversations. The conversations were split into 5 equal parts, and the occurrence of the silences was analyzed in these equal parts. Significant difference was found between the parts of the conversations in the number of silences per minute [$F(4, 76) = 3.684$, $p < 0.05$]. The least silences were found in the middle of the conversations ($mean = 1.40$ item/min), while the most of the silences occurred at the last two sections of the conversations ($mean = 2.59$ item/min, Figure 4).

The duration of the shared silences was analyzed with regard to their position of the conversation (in the 5 equal parts). The standard deviation of the values was huge and showed great overlaps; therefore data can not show any trend (Table 1).

Pauses and turn-taking

The dynamic changes of pausing was not only analyzed with regard to the equal parts of the conversation, but with regard to the TT system as well. The frequency of the SPs was analyzed with regard to their position to the TTs (near before, near

after, further before, further after TT), and significant differences were found among the positions ($\chi^2(3) = 19.599, p < 0.001$, Figure 5).

SPs occurred less frequently near the TTs than further from TTs. The frequency of the FPs were analyzed with regard to their position to the TTs, and the data showed significant differences among the positions ($\chi^2(3) = 17.65, p < 0.001$, Figure 6); FPs were the least frequent near before TTs.

The duration of the SPs and FPs was analyzed with regard to the distance from TTs. The duration of pauses did not differ significantly near TTs compared to further position from TT.

Discussion

Dynamic changes of pausing were analyzed in triadic conversations, firstly in Hungarian. The aim was to analyze how pausing changes in conversation over time as well as in the vicinity of TTs. Based on the analysis of more than 7700 items, results corroborated the first hypothesis: the frequency SPs and FPs changed over time across conversations. Pauses were the least frequent in the first and in the last sections. However, the duration of the pauses did not change over time across conversation—it may be influenced by other factors. One of these factors may be the breathiness: we found that the SPs containing audible breathing were longer than other SPs. The frequency of pauses with regard to turn-takings was analyzed as well. The SPs were less frequent in the vicinity of turn-takings than in other positions, according to our second hypothesis. Based on an earlier study for Hungarian on the same corpus (Horváth et al., 2021), the articulation rate was found to be increased in the vicinity of turn-takings. The increasing rate with the decreasing frequency of SPs signals that the current speaker is not yielding the floor yet (“rush-through”, cf. Walker, 2010). FPs occurred the least frequently near before TTs, according to our hypothesis. The analysis of shared silences showed that their frequency changed over time, however, the difference was not significant. They occurred the least frequently in the middle section of the conversations. The silences were the most frequent in the last section of conversation marking that the participants were getting run out of the topic—the fieldworker should end the conversation. The duration of pauses was not affected by the TT system significantly, contrary to our hypotheses. Our results based on conversations add new information on the timing patterns as well as on the fluency patterns of speech, which was mainly analyzed previously in narrative speech style.

Table 1. Duration of shared silences in the 5 equal parts of the conversations.

Duration of silences in the 5 parts of the conversations (ms)		
parts of conversations	mean	SD
1	541	534
2	400	426
3	388	364
4	518	460
5	625	633

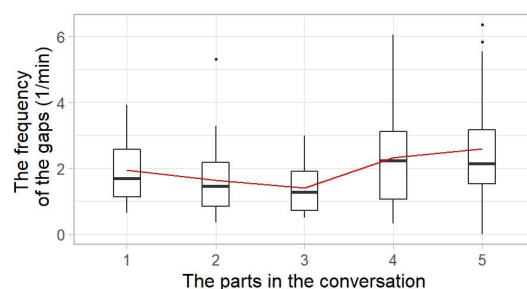


Figure 4. The frequency of shared silences in the 5 equal parts of the conversations (red line represents the means while black line on the boxes represents the medians).

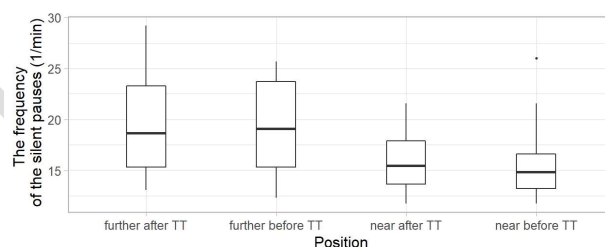


Figure 5. The frequency of the SPs according to their position to the TTs.

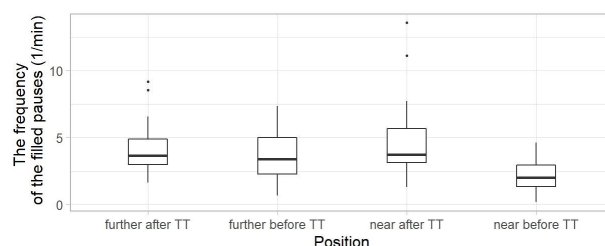


Figure 6. The frequency of the FPs according to their position to the TTs.

Acknowledgements

The research was supported by the Hungarian National Research, Development and Innovation Office of Hungary [projects No. K-128810] and the Bolyai János Research Scholarship.

References

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov & F. Caski (eds.), *Proceedings of the Second international symposium on information theory*, Budapest: Akadémiai Kiadó. 267–281.
- Bates, D., M. Mächler, B. Bolker, & S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1): 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Benus, S. 2009. Variability and stability in collaborative dialogues: Turn-taking and filled pauses. In: *Proceedings of the 10th Annual Conference of the International Speech Communication Association, INTERSPEECH*, 6–10 September, 2009, Brighton, UK, 796–799.
- Boersma, P. & D. Weenink. 2018. Praat: Doing phonetics by computer (version 6.1.38). <https://www.praat.org/> (accessed 24 January 2021).
- Brennan, S. E. & M. Williams. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language* 34(3): 383–398. <https://doi.org/10.1006/jmla.1995.1017>
- Clark, H. H. 1994. Managing problems in speaking. *Speech Communication* 15(3–4): 243–250. [https://doi.org/10.1016/0167-6393\(94\)90075-2](https://doi.org/10.1016/0167-6393(94)90075-2)
- Crible, L., L. Degand, & G. Gilquin. 2017. The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency. *Languages in Contrast* 17(1): 69–95. <https://doi.org/10.1075/lic.17.1.04cri>
- Eggins, S. & D. Slade. 1997. *Analysing casual conversation*. London, UK: Cassell.
- Horváth, V., V. Krepsz, D. Gyarmathy, Á. Hátori, J. Bóna, C. I. Dér, & Z. Weidl. 2019. Háromfős társalgások annotálása a BEA-adatbázisban: elvek és kihívások [The principles and challenges of annotating the triadic conversations in the Hungarian Spontaneous Speech Database BEA]. *Nyelvtudományi Közlemények* 115: 255–274. <https://doi.org/10.15776/NyK/2019.115.9>
- Horváth, V., V. Krepsz, D. Gyarmathy, A. Huszár, & Á. Hátori. 2021. Dynamic changes of speech patterns as cues of smooth turn-takings in Hungarian triadic conversations. Presentation at The Role of the Current Speaker in Conversational Turn Taking Workshop, 14–15 January, 2021, Berlin, Germany.
- Kuznetsova, A., P. B. Brockhoff, & R. H. B. Christensen. 2015. LmerTest Package: Tests in linear mixed effects models. *Journal of statistical software* 82(13). <http://dx.doi.org/10.18637/jss.v082.i13>
- Levelt, J. M. 1989. *Speaking. From intention to articulation*. Cambridge, MA, USA: MIT Press.
- Levinson, S. C. 1983. *Pragmatics*. Cambridge, UK: Cambridge University Press.
- Local, J. & J. Kelly. 1986. Projection and ‘silences’: Notes on phonetic and conversational structure. *Human Studies* 9(2–3): 185–204. <https://doi.org/10.1007/BF00148126>
- Neuberger T., D. Gyarmathy, T. E. Grácsi, V. Horváth, M. Gósy, & A. Beke. 2014. Development of a Large Spontaneous Speech Database of Agglutinative Hungarian Language. In: P. Sojka, A. Horák, I. Kopeček & K. Pala (eds.), *Text, Speech and Dialogue. TSD 2014, Lecture Notes in Computer Science*, Cham: Springer, 424–431. https://doi.org/10.1007/978-3-319-10816-2_51
- R Core Team 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (version 4.0.5). <https://www.R-project.org/>
- Rochet-Capellan A. & S. Fuchs. 2014. Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B* 369(1658): Article 20130399. <https://doi.org/10.1098/rstb.2013.0399>
- Sacks, H., E. Schegloff, & G. Jefferson. 1974. A simplest systematic for the organization of Turn-Taking for conversation. *Language* 50(4, part 1), 696–735. <https://doi.org/10.2307/412243>
- Schegloff, E. 1996. Turn organization: one intersection of grammar and interaction. In: E. Ochs, E. Schegloff, & S. Thompson (eds.), *Interaction and grammar*, Cambridge, UK: Cambridge University Press, 52–133. <https://doi.org/10.1017/CBO9780511620874.002>
- Stenström, A.-B. 1994. *An Introduction to Spoken Language Interaction*, London, UK: Longman.
- Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30(4): 485–496. [https://doi.org/10.1016/S0378-2166\(98\)00014-9](https://doi.org/10.1016/S0378-2166(98)00014-9)
- Torreira, F., S. Bögers, & S. C. Levinson. 2015. Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology* 6: Article 284. <https://dx.doi.org/10.3389%2Ffpsyg.2015.00284>
- Vasilescu, I., S. Rosset, & M. Adda-Decker. 2010. On the Role of Discourse Markers in Interactive Spoken Question Answering Systems. In: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, D. Tapias (eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 17–23 May, 2010, Valletta, Malta, Paper 1333.
- Walker, G. 2010. The phonetic constitution of a turn-holding practice: rush-throughs in English talk-in-interaction. In D. Barth-Weingarten, E., Reber & M. Selting, (eds.), *Prosody in Interaction. Studies in discourse and grammar*, Amsterdam, The Netherlands: John Benjamins, 51–72.

Pre-final version